ICES CM 2015/D

Assessing the transferability of biodiversity models

Yates KL[1,2], Caley MJ[1], Meeuwig JJ[3], Radford BT[4], Mellin C[1,2]

[1] Australian Institute of Marine Science, PMB 3, Townsville, QLD 4810, Australia; [2] Global Change Ecology Lab, School of Biological Sciences, University of Adelaide, South Australia 5005 Australia; [3] Centre for Marine Futures, Oceans Institute and School of Animal Biology, University of Western Australia, Crawley, Western Australia, 6009, Australia. [4] Australian Institute of Marine Science, UWA Oceans Institute, Crawley, WA 6009, Australia;

Summary

The conservation of biodiversity requires an understanding of its patterns of occurrence. In the marine environment, however, this information is often limited or completely lacking. New technologies for remotely collecting data coupled with predictive models can help address these information gaps. The utility of using remotely sensed predictors to model biodiversity metrics will depend on the transferability of models generated with biodiversity data, collected over a limited number of sites, to un-sampled areas. Here we assess the transferability of biodiversity models and the factors that affect transferability. We developed Boosted Regression Tree (BRT) models for fish species richness at each of eight sites in Western Australia. Models were used to make predictions within (internal) and between (external) locations. Distance between locations had little or no effect on model transferability. Overall, model transferability was surprisingly high, though it varied depending on the error metric used. Transferability appears to be most affected by the variability of the predictors or the response variable, suggesting that future sampling of biotic features that aims to develop models to predict into unsampled space should, where possible, focus on areas of highest abiotic and biotic variability.

Introduction

Marine biodiversity is threatened by a suite of anthropogenic impacts, many of which are increasing in distribution and/or magnitude (Halpern *et al.*, 2008). Ocean zoning and the designation of protected areas are increasingly used to conserve biodiversity and improve management of the marine environment (Douvere, 2008; Yates *et al.*, 2015). However, the scarcity of data on the distribution of marine biodiversity makes spatial planning processes less effective than they might otherwise be. One possibility for supporting more effective management is to underpin spatial planning with predictive models, whereby species and community distributions are predicted with known accuracy and precision.

Predictive models of biodiversity have advanced a great deal over recent decades, both in terms of the methods used to construct them and the technological advances in sampling methodologies. Despite these advances and associated increases in information, very substantial areas of the marine environment remain extremely poorly sampled. For many locations only abiotic and spatial data exist. It is in these situations that powerful predictive models would be of greatest utility; if a model developed for one location, where both biotic and abiotic data have been collected, can predict well the characteristics of the biotic communities at other locations where only abiotic data is available. Currently, however, it is not well understood how transferrable such models of biotic communities are, nor do we know how model transferability may be enhanced. Here we investigate the transferability of predictive models of fish species richness on the continental shelf of south Western Australia and investigate a range of factors that may affect transferability.

Materials and Methods

Using data from the Marine Futures Project (www.marinefutures.fnas.uwa.edu.au), we developed predictive models for eight locations in Western Australia. We used boosted regression tress (BRTs) (Elith *et al.*, 2008) to model the relationship between environmental characteristics, generated by acoustic sonar imagery (multibeam), and fish species richness, obtained from Baited Remote Underwater Video (BRUVS). Individual 'best' models of species richness were developed for each of the eight locations, using the combination of predictors and model parameters that resulted in the lowest mean prediction error (MPE). These models were used to make internal predictions (within the same location) and external predictions (to each of the seven other locations).

We defined transferability as external prediction performance compared to internal predictive performance; if the external prediction error was similar or lower than the internal prediction error the transferability of the model was considered to be high. We compared the predictive capacity of the internal prediction to the seven external predictions using two measures of predictive error: MPE and R^2 (predicted vs observed). We examined the change in external prediction performance (transferability) as a function of (i) increasing distance from the original model location, (ii) geographic location, and (iii) bioregion. We also measured the variability of both predictor and response variables within locations.

Results and Discussion

Model transferability depended on the metric being used. In general models developed at one location (their origin) performed worse when transferred to another location. However, the MPE of the vast majority (over 70%) of models transferred from an origin to another location fell within the range of the internal MPE for the best model at each location. Some models even performed better at other locations than at their origin. These patterns suggest that model transferability was high. However, in terms of R^2, overall model transferability was much lower and there was much more variation in predictive performance depending on the model origin and the location it was being transferred to. Increasing distance did not significantly reduce model transferability, such that model developed at distant locations had similar predictive performance as those developed neighbouring locations, but the transferability of models did vary somewhat by geographic region. Transferability of models appears to be affected by variability in either the predictors or the response variable, depending on the error metric used (MPE or R^2), with locations with low variability producing less transferable models. The transferability of future models may thus be enhanced by sampling which focuses on highly variable locations.

References

Douvere, F. 2008. The importance of marine spatial planning in advancing ecosystem-based sea use management. Marine Policy, 32: 762–771. http://linkinghub.elsevier.com/retrieve/pii/S0308597X0800064X (Accessed 11 March 2012).

Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. The Journal of animal ecology, 77: 802–13. http://www.ncbi.nlm.nih.gov/pubmed/18397250 (Accessed 25 May 2014).

Halpern, B. S., Walbridge, S., Selkoe, K. A, Kappel, C. V, Micheli, F., D'Agrosa, C., Bruno, J. F., *et al.* 2008. A global map of human impact on marine ecosystems. Science, 319: 948–52. http://www.ncbi.nlm.nih.gov/pubmed/18276889.

Yates, K. L., Schoeman, D. S., and Klein, C. J. 2015. Ocean zoning for conservation, fisheries and marine renewable energy: Assessing trade-offs and co-location opportunities. Journal of Environmental Management: 1–9. Elsevier Ltd. http://linkinghub.elsevier.com/retrieve/pii/S0301479715000614.